

Daily Report

Yuxin Ma

06.27.2013

Intro

The most valuable things today is the discussion with Prof. Tung. Some new ideas come from the discussion.

Reading

[1] was planned yesterday for it is also one of the recent papers on heterogeneous networks. This paper is a typical CHI paper that has less descriptions about the visual design and analytical procedures with much more user scenario and user study instead. The author argue that the user's exploration history is quite important in a analysis system. They have constructed a system that handles both heterogeneous network and record of user's analysis history.

Research

After discussion with Prof. Tung, there are some points that need further investigation:

- **Visualizing a model** In the DM course he mentioned that it is quite useful for both researchers and end-users to know what is happening in a “black-box” model, but not just to visualize the input and output. Take SVM as an example, the support vectors are one of the internal data generated from this kind of model. What we can do is to find a way to visualize the classification boundary based on these high-dimensional support vectors, which might be available if high-dimensional visualization techniques are applied. Also, the support vectors can represent the original data. We can construct some descirptable models like decision tree to model them again and see if there are some understandable rules that can be generated.
- **Visual Community Detection** The visual design of our enhanced parallel coordinates suffers from both data scalability and flexibility of positions of axes. Prof. Tung didn't think that the current version can deliver enough information to users with visual clutter. Also, he pointed out that in real application, users might care more about details in a community (the opinion leader or middlemen between different communities) as well as the inference (how the network and community structure evolves along time, and the cause of this evolvment). These suggestions should be considered in the future research.
- **The system** In the dinner seminar Prof. Tung mentioned their big data system and some unresolved issues. He said that after parallel mining, there is a module that combines all of the results either by combining models (such as joining many small SVMs together) or

aggregation of results. The ensemble of models is also what I was thinking about. I plan to ask him for further discussion on this point.

References

- [1] C. Dunne, N. H. Riche, B. Lee, R. Metoyer, and G. Robertson, “GraphTrail: analyzing large multivariate, heterogeneous networks while supporting exploration history,” in *CHI '12: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM Request Permissions, May 2012.